

<https://helda.helsinki.fi>

Sähköiset työkalut auttavat parantamaan terminologisen tiedon saavutettavuutta

Kudashev, Igor

Gaudeamus

2020

Kudashev , I 2020 , Sähköiset työkalut auttavat parantamaan terminologisen tiedon saavutettavuutta . julkaisussa M Hirvonen & T Kinnunen (toim) , Saavutettava viestintä : Yhteiskunnallista yhdenvertaisuutta edistämässä . Gaudeamus , Helsinki , Sivut 199-211 .

<http://hdl.handle.net/10138/328468>

unspecified

publishedVersion

Downloaded from Helda, University of Helsinki institutional repository.

This is an electronic reprint of the original article.

This reprint may differ from the original in pagination and typographic detail.

Please cite the original version.

11

SÄHKÖISET TYÖKALUT AUTTAVAT PARANTAMAAN TERMINOLOGISEN TIEDON SAAVUTETTAVUUTTA

Igor Kudashev

Termit muodostavat erikoiskielisen viestinnän ytimen. Kansainvälisen terminologia-alan järjestön TermNetin motto onkin ”Ei ole tietoa ilman terminologiaa”. Nykypäivän maailmassa tiedon määrä lisääntyy valtavan nopeasti, ja samalla myös terminologiasta tulee yhä tärkeämpää. Tässä luvussa käyn läpi, miten terminologisten kuvausten saavutettavuutta voidaan parantaa sähköisten hakuteosten, kieliteknologian ja terminhallinnan avulla.

Termeillä on keskeinen osa monikielisessä viestinnässä, muun muassa kääntämisessä. On arvioitu, että erikoisalan tiedot sisältäviä tekstejä kääntävältä kääntäjältä menee termien vastineiden selvittämiseen vähintään neljännes ajasta, pahimmillaan jopa kolme neljänneistä.¹ Tarkkaan lukuun vaikuttavat monet eri seikat, kuten erikoisala, tekstin luonne, alan tuttuus kääntäjälle sekä sanakirjojen ja muiden apuvälineiden saavutettavuus. Noin neljä viidesosaa sanakirjahauista liittyy terminologiaan,² ja jopa kaksi viidesosaa käännösvirheistä johtuu vääristä termivastineista.³ Terminologian epäjohtonmukaisuus on yleisin syy siihen, että vastaanottaja ei hyväksy käännöstä vaan vaatii sen korjaamista.⁴

Mitä parempia terminologisia vastineita käännöksessä on käytetty, sitä tarkempi käännös on, ja tämä parantaa myös käännöksen luettavuutta ja saavutettavuutta. Tarkkuudella tarkoitetaan ennen kaikkea sitä, että erikoisalaan kuuluvien käsitteiden sisältö on välitetty mahdollisimman tarkasti. Luettavuus taas edellyttää sitä, että käännöksessä käytetty tulokielinen kieliaines on lukijalle tuttua ja ymmärrettävää. Nämä tavoitteet ovat usein ristiriidassa keskenään, ja kääntäjä joutuu tasapainottelemaan niiden välillä.

Termien ymmärtäminen ja oikea käyttö vaativat toisaalta aiheen tuntemusta ja toisaalta kielellistä osaamista. Lisäksi kääntäjä tarvitsee terminologista tietoa kahdella eri kielellä. Jos omat tiedot ovat puutteellisia, joudutaan hakemaan apua hakuteoksista, kuten sanakirjoista ja termipankeista. Niistä löytyy terminologisia kuvauksia, jotka antavat tietoja termin muodosta, merkityksestä ja käytöstä. Hakuteosten tärkeä ominaisuus on tietojen tiiviys ja järjestelmällisyys.⁵

TERMI JA SANASTO KÄSITTEINÄ

Termin käsitettä on määritelty monella eri tavalla. Pohjoismaiseen suositukseen perustuvan *Terminologian sanaston*⁶ mukaan termi on erikoisalalla käytettävä yleiskäsitteen nimitys. Määritelmä rajaa pois erisnimet ja käytännössä myös yksilö- ja yleiskäsitteiden välimaastossa olevat nimikkeistöt, vaikka monilla erikoisaloilla nämä muodostavatkin ison osan erikoissanastosta. Esimerkkejä tällaisista sanoista ovat taivaankappaleiden nimet tähtitieteessä, kasvien ja eläinten lajit biologiassa ja soiden alatyypit hydrologiassa.⁷ Myös tuotenimikkeet, kuten *iPhone X* tai *Nokia 3310*, ovat epäselvässä asemassa. Termin perinteiset määritelmät eivät suoraan sulje pois esimerkiksi verbejä ja ammattislangia, mutta käytännössä näitä ilmaisukeinoja ei yleensä mielletä termeiksi, ja ne jäävät usein erikoisalan sanakirjojen ulkopuolelle.

Lisäksi kääntäjillä on usein omia käsityksiä siitä, mikä ilmaus on termi, eivätkä nämä käsitykset välttämättä ole täysin oikeita. Monille kääntäjille termit ovat ”sanoja, joita en tunne, sanakirjasta puuttuvia sanoja tai uusinta ammattislangia”.⁸ On yleinen harhaluulo, että termeillä on vain yksi merkitys ja siten ainoastaan yksi mahdollinen käännösvastine toisessa kielessä.⁹

Termin käsite on siis monimutkainen ja monitulkintainen niin teoriassa kuin käytännössäkin. Kääntämisen ja sanastotyön näkökulmasta ei kuitenkaan ole tärkeintä se, kutsutaanko jotakin kielen yksikköä termiksi vai ei, vaan se, sisällytetäänkö se terminologiseen hakuteokseen ja miten se siellä kuvataan.¹⁰ Tässä luvussa kutsun kuitenkin selvyiden vuoksi kaikkien terminkaltaisten yksiköiden kuvauksia terminologisiksi kuvauksiksi.

Myös sanastotyö-termin käyttö on Suomessa hieman ristiriitaista. Joissakin lähteissä sillä tarkoitetaan systemaattista eli järjestelmällistä terminologista työtä. Esimerkiksi *Terminologian sanasto*¹¹ määrittelee sanastotyön ja sen synonyymien terminologisen työn seuraavasti: ”työ, johon kuuluu tietyn erikoisalan käsitteitä ja niiden nimityksiä koskevan tiedon systemaattinen kerääminen, analysointi, kuvaaminen ja esittäminen”. On kuitenkin otettava huomioon, että kaikki terminologinen työ ei ole systemaattista. Kääntäjien ja tulkkien terminologinen työ tarkoittaa usein sitä, että he selvittävät tiettyyn toimeksiantoon sisältyvää termistöä. Tätä kutsutaan *ad hoc* -termityöksi eli tilapäiseksi, tiettyä tarkoitusta varten tehtäväksi terminologiseksi työksi.¹² Tällä tavalla kerättävä termistö ei useinkaan ole koko erikoisalaa ajatellen niin kattava, että sitä voitaisiin pitää järjestelmällisenä kokonaisuutena. Myöskään tällaisen terminhaku-työn menetelmät eivät useinkaan ole järjestelmällisiä, eikä työn tuloksia voida esittää systemaattisessa muodossa tai julkaista. *Ad hoc* -sanasto voidaan kuitenkin myöhemmin systematisoida ja työstää julkaistavaksi. Yleiskielessä kaikenlaisia suppeahkoja, julkaistuja tai julkaisemattomia sanakirjoja kutsutaan yleensä

sanastoiksi riippumatta siitä, onko ne laadittu systemaattisesti vai ei.

Mielestäni olisi loogista kutsua kaikenlaisten sanastojen laatimista sanastotyöksi ja puhua *systemaattisesta sanastotyöstä* silloin, kun halutaan korostaa systemaattisuutta. Sanastotyön kirjo voidaan nähdä jatkumona, jonka toisena ääripäänä on yhden termin selvittämiseksi käynnistetty *ad hoc* -selvitys ja toisena ääripäänä koko erikoisalan kattava, täysin systemaattinen sanastohanke. Näiden väliin mahtuu erikokokoisia hankkeita, joiden systemaattisuus vaihtelee.¹³ Edes Sanastokeskus TSK:n laatimat monikieliset sanastot eivät edusta systemaattisinta ääripäätä, koska ne yleensä kattavat vain tietyn erikoisalan ytimen eikä niitä laadittaessa tehdä kattavaa käsiteanalyysia tulkielessä.

NÄKÖKULMIA TERMINOLOGISTEN KUVAUSTEN SAAVUTETTAVUUTEEN

Terminologisten kuvausten saavutettavuutta voidaan tarkastella useista näkökulmista. Ensimmäinen niistä on määrällinen. Se tarkoittaa toisaalta sitä, kuinka suuri osa olemassa olevista terminologisista kuvauksista on saavutettavissa, ja toisaalta sitä, löytääkö käyttäjä kaikki tarvitsemansa tiedot terminologisista kokoelmista (kun oletetaan, että hän osaa hakea niitä oikein). On siis kysymys tarjonnan volyyymista sekä tarjonnan ja kysynnän kohtaamisesta.

Toinen näkökulma liittyy laatuun. Terminologisten kuvausten on oltava laadukkaita, ja käyttäjän täytyy myös olla varma niiden luotettavuudesta.¹⁴ Siten on tärkeää ilmaista selvästi, kuka terminologisen tiedon on hyväksynyt, milloin se on koottu ja viimeksi päivitetty ja niin edelleen. Saavutettavuus edellyttää myös sitä, että tiedot on esitetty käyttäjäystävällisessä muodossa. Lisäksi on tärkeää, että tiedot ovat saatavilla nopeasti ja juuri silloin, kun niitä tarvitaan.

Tärkeä saavutettavuuden osa-alue on myös tietojen saannin varmuus ja turvallisuus. Tähän vaikuttavia seikkoja ovat esimerkiksi tietojen pysyvyys, tekninen saavutettavuus, tietoturva ja tietosuoja. Lisäksi on huolehdittava siitä, että toiminta on kestäväällä pohjalla, eli on turvattava terminologisten kuvausten saavutettavuus myös tulevaisuudessa.

Myös palveluista tiedottaminen ja niiden integrointi muihin toimintaympäristöihin ovat erittäin tärkeitä asioita – terminologinen hakuteos on hyödytön, jos potentiaaliset käyttäjät eivät tiedä sen olemassaolosta. Lisäksi on aina syytä ottaa huomioon myös taloudellinen saavutettavuus eli se, mitä hakuteoksen käyttö maksaa.

Edellä kuvatut näkökulmat saavutettavuuteen ovat osittain päällekkäisiä ja jopa jonkin verran ristiriidassa keskenään. Esimerkiksi määrällinen ja laadullinen kehitys eivät välttämättä ole sopusoinnussa: mitä enemmän tietoja tulee saataville, sitä vaikeampi on varmistua niiden laadusta, ja myös hakunopeus ja käyttäjäystävällisyys saattavat heikentyä. Teknologia voi siis toisaalta parantaa terminologisten tietojen saavutettavuutta ja toisaalta heikentää sitä. Terminologisia hakuteoksia suunniteltaessa tulisikin huolehtia siitä, että saavutettavuuden eri osa-alueet ovat tasapainossa ja tarjottavat palvelut vastaavat kohderyhmän tai -ryhmien tarpeita mahdollisimman hyvin.

Terminologian tutkijat Gerhard Budin ja Alan Melby¹⁵ ovat analysoineet 2000-luvun alun kirjoituksessaan tärkeimpiä seikkoja, jotka heikentävät sähköisten terminologisten hakuteosten saavutettavuutta. He ovat jaotelleet nämä seikat oikeudellisiin, taloudellisiin, tiedollisiin, teknisiin ja metodologisiin esteisiin. *Oikeudellinen este* muodostuu esimerkiksi silloin, kun terminologisten hakuteosten omistajien on vaikea määritellä, millä ehdoilla he olisivat valmiita jakamaan hakuteosten sisältämiä tietoja. Koska termit välittävät informaatiota, terminologian jakamiseen sisältyy riskiä, esimerkiksi riski liikesalaisuuksien tahattomasta paljastamisesta kilpailijoille. Budin ja Melby

ovat huomauttaneet myös, että kielellisten resurssien tekijänoikeuksiin liittyviä kysymyksiä ei ole yksiselitteisesti ratkaistu. *Taloudelliset esteet* taas viittaavat siihen, että terminologisten hakuteosten käyttö on potentiaalisille käyttäjille liian kallista. *Tiedolliset esteet* tarkoittavat sitä, että potentiaaliset käyttäjät eivät välttämättä edes tiedä terminologisen hakuteosten olemassaolosta. *Tekniset esteet* voivat olla monenlaisia. Budin ja Melby mainitsevat muun muassa standardoimattomien tietoformaattien käytön sekä tarvittavien työkalujen, kuten terminhallintaohjelmien ja konvertointityökalujen, puuttumisen. *Metodologiset esteet* ilmenevät muun muassa siinä, että terminologisten hakuteosten laatimisessa käytetään erilaisia menetelmiä ja malleja, minkä takia ne eivät ole tasalaatuisia ja yhteensopivia.

Erilaisten sähköisten työkalujen ansiosta terminologisia tietoja voidaan myös räätälöidä käyttäjän ja käyttötilanteen mukaan. Vaikka tästä aiheesta on vasta hyvin vähän tutkimustietoa,¹⁶ voidaan pohtia, mikä on ollut teknologian vaikutus terminologisen tiedon saavutettavuuteen tähän mennessä ja mitä ratkaisuja saavutettavuuden parantamiseksi voidaan tarjota lähitulevaisuudessa. Seuraavassa alaluvussa käsittelemme tätä aihetta analysoimalla 2000- ja 2010-luvuilla käynnistettyjä suomalaisia ja eurooppalaisia hankkeita, jotka koskevat terminhallintaa.

TERMINHALLINTAAN LIITTYVIÄ HANKKEITA

Terminologisten hakuteosten hakemistot

Terminologisen hakuteosten saavutettavuus edellyttää, että potentiaaliset käyttäjät ovat tietoisia hakuteoksen olemassaolosta ja käyttöehdoista. Jos tietoja on monissa eri lähteissä, niiden löytämistä helpottaa terminologisten hakuteosten hakemisto. Tällaisia hakemistoja löytyy internetistä paljon, mutta

niiden kattavuus ja luotettavuus vaihtelevat paljon. Suurin osa hakemistoista on pelkkiä linkkikokoelmia, jotka keskittyvät sangen rajallisiin aihepiireihin ja kielivalikoimiin. On kuitenkin myös kattavia hakemistoja, joissa on esimerkiksi jonkinlainen systemaattinen aihealueuokittelu. Yksi esimerkki tällaisesta hakemistosta on Vaasan yliopiston professori Anita Nuoppo- sen laatima ja ylläpitämä TERM-ONLINE-linkkikokoelma.

Kehittynein hakemisto lienee tällä hetkellä Euroopan par- lamentin terminologiaa koordinoivan TermCoord-yksikön ylläpitämä Glossary Links. Se sisältää linkkejä yli 5 000 online- sanastoon, ja sieltä voi hakea sanastoja aihealueen, sanaston otsikossa esiintyvän avainsanan sekä kielen perusteella. Term- Coord tarkastaa sanastojen luotettavuuden ja päivittää linkkejä säännöllisesti. Vastaavankaltaisen kattavan, jatkuvasti päivitet- tävän hakemiston perustaminen myös Suomessa käytettävistä online-sanastoista olisi hyödyllistä.

Kansalliset termipankit

Euroopassa on 1990-luvun jälkeen perustettu yli kaksikym- mentä kansallista termipankkia. Tällainen on esimerkiksi Ruotsin Rikstermbanken. Kansallista termipankkia suunniteltiin myös Suomessa, mutta Kansalliskielistrategian vuoden 2017 toimintasuunnitelmassa kansallista termipankkia ei enää mainita. Jos kansallinen termipankki on suunniteltu asian- mukaisesti ja sen tavoitteet ja budjetti ovat realistisia, se voi parantaa terminologisten kuvausten saavutettavuutta ratkai- sevasti. Ihannetapauksessahan terminologisia kuvauksia saa yhdestä paikasta ja ne ovat laadukkaita, luotettavia ja ilmaisia.

Kansallisen termipankin toteuttamiseen sisältyy tiettyjä käytännön ongelmia. Eri lähteistä tulevia, eri lailla laadittuja aineistoja voi olla vaikea sovittaa yhteen. Samat käsitteet voi- vat olla määriteltyinä monista eri näkökulmista, eikä ole selvää, minkä tahon määritelmä tulee valita päämääritelmäksi ja mitä

muille määritelmille silloin tehdään. Suuria haasteita tuovat myös puuttuvat lähdeviittaukset ja vajanaiset hallinnolliset tiedot (esimerkiksi tieto siitä, mistä kuvaukset ovat peräisin ja milloin ne on viimeksi päivitetty). Näitä ongelmia on puitu terminologia-alan konferensseissa, kuten NordTerm- ja TKE-konferenssien (TKE, Terminology and Knowledge Engineering) yhteydessä pidetyissä työpajoissa sekä Euroopan terminologiajärjestö EAFT:n huippukokouksen yhteydessä.

Metodologiset ja tekniset ongelmat ovat kuitenkin ratkaistavissa, kunhan niihin paneudutaan hyvin. Valitettavasti kansallisten termipankkien suurimpana ongelmana kaikkialla Euroopassa on ollut rahoituksen riittämättömyys, epävarmuus ja pirstaleisuus. Kansallinen termipankki ei ole muutaman vuoden projekti, vaan se vaatii jatkuvaa ylläpitoa ja kehittämistä eli pysyvää henkilöstöä.

Euroopan interaktiivinen termipankki IATE

IATE (Interactive Terminology for Europe) on Euroopan unionin elinten käännöskeskukseen perustama maailman suurin termitietokanta. Se sisältää lähes miljoona termitietuetta eli lähes kahdeksan miljoonaa termiä yli 20 kielellä. IATEsta löytyy ennen kaikkea Euroopan unionia koskeva termistö mutta myös keskeistä termistöä sellaisilta aloilta kuin oikeustiede, maatalous ja tietotekniikka.¹⁷

IATEn tarjonnassa on suuria eroja eri kielten välillä. Suurimpia kieliä ovat englanti (yli 1,3 miljoonaa termiä), ranska (hieman alle 1,3 miljoonaa termiä) ja saksa (lähes miljoona termiä). Suomi sijoittuu termien määrässä hieman keskitason alapuolelle (noin 330 000 termiä).¹⁸

IATEn uusimmassa versiossa (2018) on tehty muutoksia tietueiden luomiseen ja muokkaamiseen, viitteiden hallintaan, aihealueen valitsemiseen, käyttäjien hallintaan, muutoshistorian hallintaan, ja lisäksi se on entistä tiiviimmin ja

monipuolisemmin integroitavissa tietokoneavusteisen kääntämisen työvälineisiin. Lisäksi IATEsta on tarkoitus poistaa vanhentuneet ja epäluotettavat tiedot.¹⁹ Uudistettukaan IATE ei silti pysty täysin korvaamaan kansallisia termipankkeja, koska se perustuu koko Euroopan unionin näkökulmaan. IATE on kuitenkin erittäin arvokas terminologinen resurssi ja uranuurtaja terminhallinnan metodologiassa ja teknologiassa.

EuroTermBank ja Terminology as a Service -hankkeet

Keskitettyjen termipankkien lisäksi on olemassa niin sanottuja federoituja termipankkeja, joissa hallitaan yhteisen teknologian avulla useaa erillistä sanastoa. Käyttäjälle tällaiset palvelut näyttäytyvät käyttöliittymänä, jonka kautta pystyy hakemaan tietoja useista lähteistä niin, että nämä lähteet erottuvat kuitenkin erillisinä ja itsenäisinä kokonaisuuksina. Federointia on toteutettu kahdessa isossa Euroopan unionin hankkeessa, jotka ovat EuroTermBank²⁰ ja Terminology as a Service (TaaS).²¹

EuroTermBank-hankkeessa (2005–2006) tavoitteena oli yhdistää erityisesti Baltian maiden, Puolan ja Unkarin terminologisia hakuteoksia. Tuloksena syntyi suurehko, monikielinen federoitu termipankki. Myöhemmin sen toiminnallisuutta laajennettiin muun muassa pilvipalveluiden avulla, ja myös sen liittäminen erilaisiin työkaluihin, kuten memoQ-käännösohjelmaan, tuli mahdolliseksi. EuroTermBankissa on tällä hetkellä yli kaksi miljoonaa termiä 33 kielellä, ja se on linkitetty neljään ulkopuoliseen tietokantaan.

TaaS-hankkeessa (2012–2014) huomion kohteina olivat kieliteknologiat, kuten termien ja vastineiden automaattinen löytäminen, sekä työkalut terminologian jakamiseen ja uudelleenkäyttöön. Työn pohjana käytettiin olemassa olevia resursseja, kuten IATEa, EuroTermBankia ja TAUS Dataa.²² TaaS-hanke oli erittäin kunnianhimoinen ja teknologian kehittämisen näkökulmasta hyvin onnistunut, varsinkin kun otetaan

huomioon sen lyhyt toteutusaika. Toisaalta se osoitti, että jos automaattiset teknologiat yhdistetään valtaviiin, keskenään linkitettyihin termikokoelmiin, toistojen ja epärelevanttien tietojen määrä kasvaa sietämättömäksi.

Tieteen termipankki

Suomi on ollut edelläkävijä yhteisöllisessä termityössä. Konkreettinen, laajamittainen esimerkki tällaisesta työstä on Suomen Akatemian rahoittama Tieteen kansallinen termipankki -infrastruktuurihanke (2011–2015). Tieteen kansallinen termipankki (josta käytetään yleisesti nimitystä Tieteen termipankki, TTP) rakentaa kaikkien Suomessa harjoitettavien tieteenalojen yhteisen, avoimen ja jatkuvasti päivitettävän termitietokannan tiedeyhteisön ja kansalaisten käyttöön. Vaikka Tieteen termipankki toimii wikialustalla, joukkoistaminen on rajoitettu kunkin tieteenalan asiantuntijoihin: tutkijayhteisö ottaa itse vastuun oman erikoisalansa sanaston tuottamisesta verkkopalveluun sekä tietojen täydentämisestä ja päivittämisestä.²³ Tätä kirjoittaessani termipankki sisälsi yli 150 000 erikielistä termiä 41 erikoisalalta. Termien ohien on useimmiten lisätty alan asiantuntijoiden laatimat määritelmät ja selitteet.

TEPA

Sanastokeskus TSK:n ylläpitämä TEPA (”Tekniikan termipankki”) on Suomen vanhin avoin termipankki.²⁴ Sen rungon muodostaa TSK:n oma aineisto, joka on suurimmaksi osaksi laadittu systemaattisen sanastotyön periaatteita noudattaen ja on erittäin laadukas, joskin osittain vanhentunut. TSK:n oma osuus sisältää erikoisalojen termejä, määritelmiä, selitteitä ja käsitelkaavioita sekä vieraskielisiä vastineita.

TSK:n oman aineiston lisäksi TEPA sisältää muiden tahojen, esimerkiksi ruotsalaisen Terminologicentrum TNC:n,

Kotimaisten kielten keskuksen, IATEn ja Tieteen termipankin, laatimia aineistoja. Ulkopuolisten tuottama sisältö on TEPA-termipankissa niin sanottuna paikallisena kopiona, mutta lisäksi tarjotaan linkit alkuperäiseen tietueeseen. Tällainen ”offline-federointi” on parantanut hakunopeutta ja kasvattanut saatavilla olevien terminologisten kuvausten määrää, mutta samalla aineistosta on tullut selvästi heterogeenisempi. Jos termillä on monta merkitystä (ks. esimerkiksi *palo* tai *kieli*), sopivan käsitteen erottaminen hakutuloksista vie aikaa. Lisäksi TEPAan integroidusta IATEn aineistosta puuttuvat määritelmät ja selitteet. Näistä puutteista huolimatta TEPA on vahvasti matkalla kohti kansallisen termipankin asemaa.

LOPUKSI

Terminologisten tietojen saavutettavuutta voidaan tarkastella useista näkökulmista, jotka liittyvät muun muassa määrään, laatuun, tietojen saannin nopeuteen ja varmuuteen, turvallisuuteen sekä hinnoitteluun. Nämä näkökulmat ovat osittain päällekkäisiä ja kilpailevat toistensa kanssa, joten yhtä osa-aluetta kehitettäessä muut voivat heikentyä. Saavutettavuudelle on myös olemassa monenlaisia esteitä, esimerkiksi informatiivisia, oikeudellisia, taloudellisia, teknisiä ja metodologisia seikkoja. Terminologisia hakuteoksia suunniteltaessa tulisikin kiinnittää huomiota kaikkiin näihin seikkoihin ja huolehtia siitä, että saavutettavuuden eri osa-alueet olisivat tasapainossa ja vastaisivat kohderyhmän tai ryhmien tarpeita mahdollisimman hyvin.

Yhtenä terminhallinnan megatrendinä on 2000-luvulla ollut tarve koota terminologiset tiedot yhteen paikkaan tai ainakin tarjota terminologisia tietoja ”yhden luukun periaatteella”. Esimerkkejä tällaisista palveluista ovat useat kansalliset termipankit ja Euroopan unionin IATE. Niin sanottu federointi voi olla keskitettyä termipankkia joustavampi vaihtoehto, mutta aito,

reaaliaikainen federointi näyttää olevan haasteellista. Monet termipankit, kuten EuroTermBank ja TEPA, ovatkin ”hybridimalleja”, eli ne tarjoavat sisältöä ulkopuolisista tietokannoista paikallisina kopioina.

Suurimpien termipankkien koko on nykyään miljoonia termejä. Termitietueiden lisääntyvään määrään sisältyy kuitenkin ongelmia: mukaan voi jäädä vanhentunutta ja epäluotettavaa aineistoa sekä paljon toistoa, aineisto on heterogeenista eikä kaikille tietueille anneta hallinnollisia tietoja. Relevantteja tietoja on myös vaikeampi erottaa tietotulvasta. Kieliteknologiasta voi tulevaisuudessa olla suurta hyötyä esimerkiksi kaksoiskappaleiden automaattisessa poistamisessa sekä tietojen suodattamisessa hakijan määrittelemien kriteerien avulla. Myös laadunvarmistuksen tärkeys korostuu entisestään.

Terminologisten kuvausten tuottaminen vaikuttaa ratkaisevasti terminologisen tiedon saavutettavuuteen. On kuitenkin hankalaa varmistaa, että tähän osa-alueeseen panostetaan riittävästi. Monet eri tahot, kuten yksittäiset yritykset, ovat kyllä kiinnostuneita yhdistämään ja hallitsemaan valmiita sisältöjä, koska se lisää tuottavuutta, mutta sisällöntuottaminen ei niitä kiinnosta. Myöskään julkisista varoista ei tueta sisällöntuottamista riittävästi. Näihin haasteisiin on yritetty vastata siirtymällä yhteisölliseen sisällöntuottamiseen, jakamiseen ja validointiin (esimerkiksi Tieteen termipankki ja TaaS). Kokemus on kuitenkin osoittanut, että riittävän laadukkaan lopputuloksen saamiseksi tarvitaan termityön asiantuntijoiden hallinnoimaa, rajatun joukon toteuttamaa joukkoistustyötä.

Kieliteknologian ja terminhallinnan yhdistäminen on toistaiseksi ollut melko vähäistä ja varovaista. Viime aikoina on esimerkiksi alettu integroida terminhallintajärjestelmiä automaattisen termi- ja vastinepoiminnan sekä käännöstyökalujen kanssa (IATE ja TaaS). Sähköinen ympäristö ja kieliteknologia mahdollistavat myös sisällön räätälöinnin käyttäjän ja käyttötilanteen mukaan. Tästä aiheesta on kuitenkin toistaiseksi vain

vähän tutkimustietoa. Myöskään kattavaa teoriaa terminhallinnasta sekä terminhallinnan ja kieliteknologian vuorovaikutuksesta ei ole, ja sitä kaivataan kipeästi.

Kieliteknologialla on varmasti keskeinen osa tulevaisuuden termipankeissa. On kuitenkin tärkeää ymmärtää, että isoja termipankkeja ei voida kehittää projektinomaisesti vaan tarvitaan pysyvä rahoitus ja asiantunteva, sitoutunut henkilöstö. Tämä sijoitus maksaisi itsensä varmasti takaisin.

**11. SÄHKÖISET TYÖKALUT
AUTTAVAT PARANTAMAAN
TERMINOLOGISEN TIEDON
SAAVUTETTAVUUTTA**

1. Ubin 1992, 16; Grinev 1993, 259.
2. Ks. Tomaszczyk 1989, 177.
3. Ubin 1992, 15.
4. SDL 2016, 7.
5. Ks. Bergenholtz & Tarp 1995,
15–16; Kudashev 2013b, 24–27;
2007, 22–47.
6. *Terminologian sanasto* 2006, 22.
7. Ks. esim. Kudashev 2005.
8. Bononno 2000, 648.
9. Miram 2001, 62–63.
10. Gerd 1986, 36–37.
11. 2000, 31.
12. Wright 1997, 19; Wright & Wright
1997, 147; Bononno 2000, 646.
13. Vrt. tämän teoksen luku 15.
14. Vrt. Varantola 1998, 14; Nkwenti-
Azeh 2001, 604–606.
15. Ks. Budin & Melby 2000.
16. Trap-Jensen 2010, 1134.
17. About IATE 2019; Peltoniemi
2017.
18. Peltoniemi 2017.
19. Mt.; New developments 2017;
IATE Brochure 2019.
20. Ks. Rirdance & Vasiljevs 2006;
Vasiljevs 2011.
21. TaaS 2014.
22. Mt., 5.
23. Onikki-Rantajääskö & Pitkänen-
Heikkilä 2015, 15.
24. Ylisalmi 2018.